Transportation Research Interdisciplinary Perspectives 39 (2026) 102223

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.elsevier.com/locate/trip

Trust in automated mobility as a social process: Behavioural and conversational evidence from shared rides

Ugnė Aleksandra Morkutė [a,*], Pavlo Bazilinskyy [b], Jom Dieben [c], Francesco Walker [c]

[a] New York University, Tandon School of Engineering, 6 MetroTech Center, Brooklyn, NY 11201, United States
[b] Eindhoven University of Technology, Eindhoven, The Netherlands
[c] Leiden University, Institute of Psychology, Leiden, The Netherlands

ARTICLE INFO

Dataset link: https://osf.io/7d4mw

Keywords:
Trust in automation
Automated vehicles
Group dynamics
Human–automation interaction
Virtual reality

ABSTRACT

Trust is a critical determinant of public acceptance of automated vehicles, particularly in shared mobility contexts where passengers co-experience automation. While prior research has largely examined trust as an individual judgement, emerging shared automated mobility raises the need to understand how trust unfolds within groups and across different sources of driving agency. This study investigates how situational urgency and driving agency (human vs. automated control) shape trust during shared rides. Thirty-six participants completed a within-subjects virtual reality (VR) study in triads, experiencing automated and manual driving under low- and high-urgency scenarios. Trust was assessed using a multi-method approach combining real-time behavioural signals (press-and-hold trust-loss input), post-trial questionnaires, and thematic analysis of group conversations. Results show that trust was lowest in automated low-urgency scenarios and highest in manual high-urgency scenarios, suggesting that perceived agency strongly shapes trust calibration. Button-press behaviour captured moment-to-moment trust breakdowns and correlated with subjective trust ratings, supporting its validity as a real-time indicator. Qualitative findings reveal that trust was socially situated, with group sense-making processes shaping how passengers interpreted vehicle behaviour, situational ambiguity, and control attribution. Together, the findings position trust in automated mobility as a socially embedded and context-dependent phenomenon, and highlight the importance of designing shared automated vehicle (AV) systems that support real-time trust expression and collective understanding.

1. Introduction

The emergence of automated vehicles (AVs) marks a significant development in contemporary transportation systems, with the potential to improve road safety and enhance mobility efficiency. A substantial proportion of traffic crashes are attributed to human error, with estimates exceeding 94% (National Highway Traffic Safety Administration, 2008; Tafidis et al., 2022; Tengilimoglu et al., 2023). By reducing human involvement in complex driving tasks, AVs are expected to enable smoother traffic flow (Bagloee et al., 2016), greater compliance with traffic regulations (Fagnant and Kockelman, 2015), and increased in-vehicle productivity through support for non-driving-related tasks (NDRTs) (Biondi et al., 2019; Pettersson and Karlsson, 2015; Wilson et al., 2022). Despite these anticipated benefits, public trust in AVs remains limited (Bazilinskyy et al., 2015), particularly in situations characterised by uncertainty or elevated risk (Shepardson, 2024).

Existing research on trust in AVs has predominantly examined individual passenger judgements and behavioural responses to automation (Choi and Ji, 2015; Hoff and Bashir, 2015; De Cet et al., 2024). However, as shared automated vehicles (SAVs) become increasingly plausible, understanding how trust is shaped within groups of co-present passengers is gaining importance. Shared rides create opportunities for social reassurance, verbal negotiation, and collective interpretation of driving events, introducing interpersonal dynamics that may shape trust formation. These interpersonal processes may substantially influence how trust develops, yet remain underexplored in the context of automated mobility (Momen et al., 2023). Furthermore, previous work has not combined group-level conversational analysis with real-time behavioural markers of trust during simulated SAV journeys involving triads of passengers. This combination is theoretically important for three reasons. First, triads provide a minimal yet meaningful social unit in which interpersonal sense-making can emerge while remaining analytically tractable. Second, immersive simulation enables controlled manipulation of urgency and driving agency that would be difficult and unsafe to reproduce in real-world settings. Third, integrating real-time behavioural signals with conversational

* Corresponding author.
E-mail address: uam2012@nyu.edu (U.A. Morkutė).

https://doi.org/10.1016/j.trip.2026.102223
Received 24 March 2026; Received in revised form 22 August 2026; Accepted 25 August 2026

Abbreviations

AV	Automated vehicle
SAV	Shared automated vehicle
VR	Virtual reality
NDRT	Non-driving-related task
OBS	Open Broadcaster Software
RQ	Research question
CI	Confidence interval

data allows trust to be captured both as an immediate experiential response and as a socially negotiated interpretation. Together, this design enables a more process-oriented account of trust in automated mobility as a socially embedded and temporally unfolding phenomenon rather than a purely individual judgement.

The present study investigates how situational urgency, driving agency (human vs. automated control), and interpersonal interaction shape trust during shared rides in an AV. Manual driving conditions were included as a comparative baseline to examine how trust calibration changes when responsibility for driving is attributed to a human versus an automated system within the same shared mobility context. Thirty-six participants completed a within-subjects virtual reality (VR) driving simulation in groups of three, experiencing automated and manual driving conditions under low- and high-urgency scenarios. Trust was assessed using three complementary approaches: (1) real-time button-press behaviour indicating moment-to-moment changes in perceived safety, (2) post-trial questionnaires capturing reflective evaluations, and (3) thematic analysis of group conversations. Together, these methods provide an integrative perspective on how passengers experience, express, and socially negotiate trust during shared automated mobility.

Building on this motivation, this article makes three contributions. First, conceptually, it frames trust in automated mobility as a socially embedded process that unfolds through real-time group interaction rather than purely individual judgement. It further introduces passengers' sense of individual autonomy as an underexplored dimension shaping trust in shared rides — one that connects perceived driving agency to the social channels through which passengers retain or lose felt influence over the journey. Second, methodologically, it introduces a multi-method approach that integrates real-time behavioural input (press-and-hold trust-loss signalling) with questionnaires and conversational analysis to capture trust dynamics across temporal and social levels. Third, empirically, it shows that trust was lowest in automated low-urgency scenarios and highest in manual high-urgency scenarios, highlighting how perceived driving agency and situational context shape trust calibration.

These questions are particularly relevant as shared automated mobility, such as robotaxis and autonomous ride-sharing services, becomes more plausible and shifts attention from individual to collective experiences of automation. Importantly, the primary goal of this study is to examine trust as a dynamic and socially embedded process in shared mobility contexts. The use of multiple modalities (behavioural, subjective, and conversational) is intended to capture this process at different temporal and social levels, rather than to compare measurement approaches. In this sense, these measures are treated as complementary lenses on trust formation rather than as competing methods.

2. Background

2.1. Trust in automated vehicles

Trust in automation is shaped by dispositional, situational, and learned factors (Hoff and Bashir, 2015; Lee and See, 2004; Walker et al., 2023). In the context of AVs, studies have examined how perceived reliability, predictability, transparency, and user experience influence the willingness to rely on automated systems (Choi and Ji, 2015; Walker et al., 2023). Trust is commonly conceptualized as the belief that an automated system will act in ways that support user goals under uncertainty (Lee and See, 2004). It is understood as dynamic and subject to recalibration based on system performance (Hoff and Bashir, 2015; Muir, 1994). Achieving calibrated trust—neither over-reliance nor unwarranted distrust—is critical for safe and effective interaction with AVs (Walker et al., 2023; Kaufman et al., 2025).

A related but underexplored dimension of trust concerns passengers' sense of individual autonomy — the subjective experience of retaining influence or the capacity to intervene in a journey. Autonomy is identified as a basic psychological need shaping motivation and wellbeing (Ryan and Deci, 2000), and its relevance extends to human–machine interaction: when users perceive they lack voice or control over a system's actions, negative affect and reduced reliance often follow. In shared automated mobility, passengers may experience a reduction in felt autonomy not only because control is delegated to the system, but because the social channels through which they might otherwise influence a human driver — verbal communication, shared reactions, mutual monitoring — are largely unavailable. Despite its theoretical relevance, sense of individual autonomy has received limited direct attention in the AV trust literature, representing a gap this study begins to address empirically.

2.2. Measuring trust: From self-report to real-time indicators

Trust in AVs has traditionally been assessed using questionnaires (Choi and Ji, 2015; Dikmen and Burns, 2017; Weinstock et al., 2012; Nordhoff et al., 2021), which provide valuable but retrospective insight. Increasingly, researchers highlight the need for real-time behavioural measures of trust (Lu and Sarter, 2019; Ayoub et al., 2023). Gaze behaviour has been widely explored, with higher trust often associated with reduced road monitoring and increased engagement in NDRTs (Hergeth et al., 2016; Walker et al., 2019; Felbel et al., 2022), and lower trust linked to vigilant scanning and physiological stress (Strauch et al., 2019; Walker et al., 2019). Physiological signals, such as electrodermal activity (EDA), respiration, and heart rate, have also been used to infer trust and user state (Meteier et al., 2023). Neuroimaging approaches have also been explored: functional near-infrared spectroscopy (fNIRS) has been used to objectively measure trust in highly automated driving, revealing distinct cortical signatures for trust and distrust (Perello-March et al., 2023). However, gaze–trust correlations remain inconsistent (Gold et al., 2015), and a systematic review of 258 empirical studies found that existing methodologies are predominantly experimental and individual-focused, calling for frameworks that integrate contextual and ethnographic approaches to capture trust beyond momentary human–machine interaction (Raats et al., 2020).

The present study incorporates a real-time button-press mechanism, allowing participants to signal moments of reduced trust during driving. This simple but unobtrusive input offers a promising complement to questionnaire-based assessments, enabling fine-grained observation of trust fluctuations as they occur, while group conversational analysis extends this approach to the shared social context in which trust unfolds.

2.3. Trust in context: Situational urgency

Trust in AVs is shaped not only by system performance but also by contextual factors (Hoff and Bashir, 2015). In low-urgency scenarios, consistent and predictable behaviour can reinforce trust over time (Sun et al., 2020). High-urgency situations, including sudden hazards, unclear feedback, or system hesitation, pose greater challenges, potentially amplifying stress and eroding confidence (Thomaschke and Haering, 2014; Atakishiyev et al., 2024). Perceptions of safety depend on the assertiveness, smoothness, and transparency of AV responses (Fu et al., 2020; Taylor et al., 2023; Ju et al., 2022). Although trust may be restored following corrective action (Kraus et al., 2020; Baccari et al., 2024; Hossain et al., 2019), empirical understanding of how trust evolves during and after high-urgency events remains limited.

2.4. Trust as a social process

Shared mobility settings introduce interpersonal dynamics that shape passenger perception. Social proof suggests that people look to others' reactions for guidance during uncertainty (Venema et al., 2020). In SAVs, conversational cues—such as explicit discussion, reassurance, or expressions of doubt—can influence trust judgements (Momen et al., 2023). Passengers may respond not only to the vehicle but to each other's interpretations, comfort levels, and behavioural signals (Yoo et al., 2024; Dolins et al., 2021). In some cases, anxiety about co-passengers can outweigh concerns about the vehicle itself (Dolins et al., 2021).

Group-level trust dynamics have been explored in adjacent domains: research on interactive team cognition demonstrates that collective sense-making resides in team interactions rather than individual knowledge (Cooke, 2015), and recent work on multi-human–AI teams shows that trust in AI can spread between team members through verbal and non-verbal cues—a phenomenon termed trust contagion (Rojas et al., 2025). These findings suggest that interpersonal dynamics fundamentally shape how trust in automated agents is formed and calibrated within groups, yet such processes remain underexplored in shared automated vehicle contexts.

This social dimension of trust has broader adoption implications: a meta-analysis of 72 AV acceptance studies (N = 32,540) identified social influence as one of the strongest determinants of adoption intentions alongside trust and perceived safety (Adnan, 2024), suggesting that understanding how trust is collectively shaped in shared rides has direct relevance for uptake at the population level. Recent qualitative work similarly identifies AI-scepticism and shifting human–machine interdependence as under-researched psychosocial barriers to AV acceptance that remain poorly captured by quantitative adoption models (Miskolczi et al., 2025), dimensions that group conversations in shared rides may surface in real time. We therefore conceptualize the processes operating within shared rides as socially embedded trust: an emergent, dynamic phenomenon unfolding within real-time interpersonal coordination.

2.5. Aim of the study

This study examines trust in AVs as a dynamic, context-sensitive, and socially embedded process. Moving beyond individual assessments, it investigates how trust is experienced and expressed in shared rides and how it fluctuates in response to situational urgency, driving agency (human vs. automated control), and interpersonal interaction. The following research questions guided the study:

1. How does trust fluctuate in response to situational urgency and driving agency during shared rides?
2. Can button-press behaviour function as a real-time behavioural indicator of trust loss?
3. How is trust in AV performance co-constructed through group conversation during shared rides?

Manual driving trials were included as a theoretical baseline to examine how perceived agency shapes trust calibration. Rather than treating manual driving as a non-AV condition, this comparison allowed us to isolate how trust judgements shift when responsibility for driving is attributed to a human versus an automated system within the same shared ride context.

3. Method

3.1. Participants

A power analysis using G*Power 3 (Faul et al., 2007) indicated that 36 participants were required to achieve 95% power for a repeated-measures design with a medium effect size (f = 0.25, d = 0.5, α = 0.05). Participants (aged 18–30) were recruited over 12.5 weeks via mixed convenience sampling, primarily from university student populations. Some participants enrolled individually, while others registered together with friends or acquaintances. As a result, triads varied in prior familiarity, ranging from fully acquainted groups to mixed or unfamiliar compositions. This variability was not experimentally controlled but reflects realistic shared mobility contexts, where co-passengers may or may not know each other. We acknowledge that prior familiarity may influence conversational dynamics and socially embedded trust formation, and return to this point in the Limitations.

Eligibility criteria included holding a valid driver's licence, English proficiency, normal or corrected-to-normal vision, and no prior experience as a passenger in an automated vehicle. Participants received two research credits as compensation. Ethical approval for the study was obtained from the Psychology Research Ethics Committee of the [anonymous university].

3.2. Apparatus

The experimental setup consisted of three Next Level Racing seats equipped with Logitech driving hardware and Varjo headsets with integrated eye tracking. The simulation was rendered in high-fidelity VR using Unity and the Coupled Simulator, an open-source platform designed to support multi-user research in traffic environments (Bazilinskyy et al., 2020). Button-press input was collected via handheld VR controllers. Group conversations were captured using a Blue Snowball microphone and recorded with Open Broadcaster Software (OBS).

3.3. Materials and stimuli

The driving environment consisted of a realistic urban–highway setting presented in VR. Situational urgency was manipulated through route-based events. In the low-urgency scenario, participants encountered a routine exit from a highway back into the city (Fig. 1). In the high-urgency scenario, they encountered two critical events: a roadblock requiring an evasive manoeuvre and a flipped car obstructing part of the lane (Fig. 2). These events were selected to establish a clear contrast in perceived risk and time pressure.

To support construct validity, the urgency manipulation was designed to introduce clear differences in perceived time pressure and situational risk between conditions. Low-urgency trials involved routine navigation without immediate hazards, whereas high-urgency trials included sudden obstacles requiring rapid interpretation and response. No explicit self-report manipulation check was collected.

3.4. Measures

In this study, trust was operationalized as confidence in the entity responsible for driving decisions within each condition. In automated trials, participants evaluated trust in the automated driving system as the primary control agent. In manual trials, trust judgements primarily concerned the human driver as the active decision-maker. This conditional framing ensured that trust always referred to the agent perceived to be in control of vehicle behaviour, rather than to the vehicle or social group in general. Importantly, while conversations reflected broader social dynamics, all quantitative trust measures (questionnaires and button presses) were anchored to the currently active driving agent. This operationalization allowed us to examine how trust calibrates as perceived responsibility shifts between human and automated control within a shared mobility context.

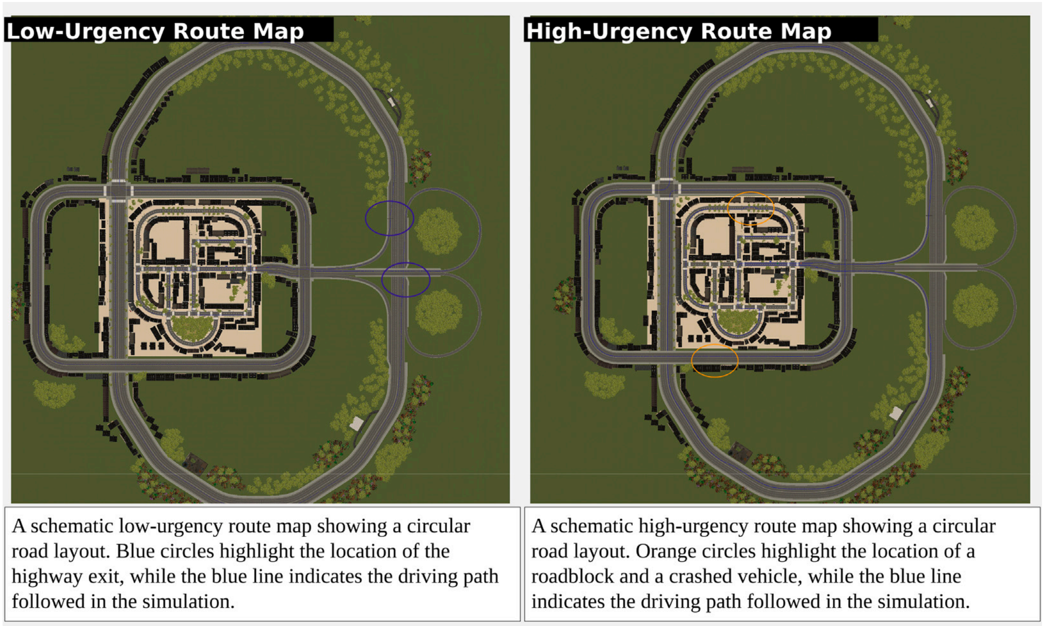

Fig. 1. Schematic route maps for the low- and high-urgency scenarios. Blue circles indicate highway exits (low urgency). Orange circles indicate roadblocks and a flipped car (high urgency).
Note. The maps illustrate how situational urgency was manipulated using specific road events distributed along fixed routes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

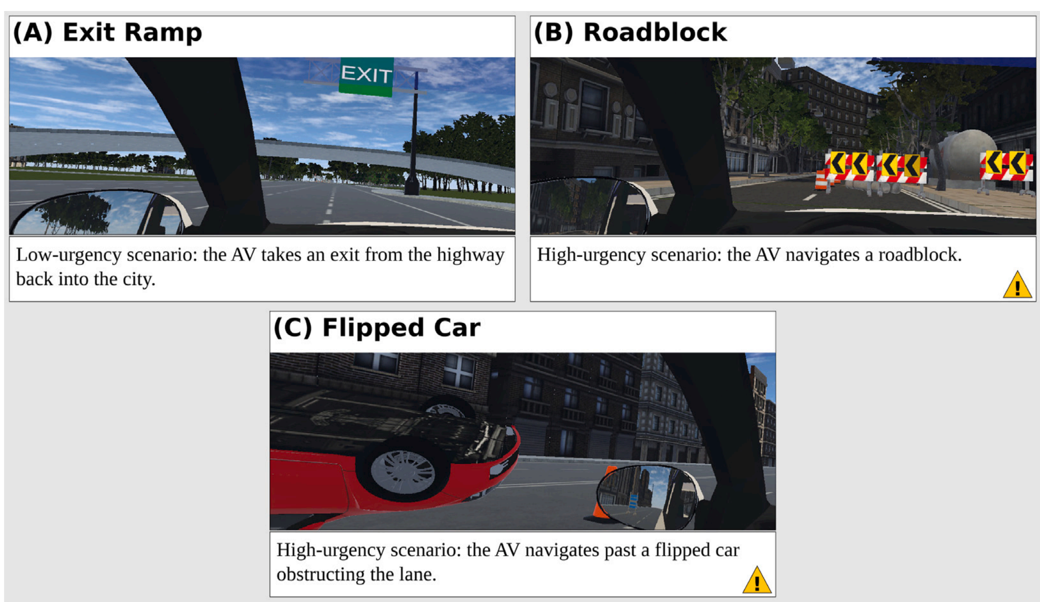

Fig. 2. Screenshot examples of low- and high-urgency events: (A) Exit ramp (low urgency), (B) roadblock (high urgency), and (C) flipped car (high urgency).
Note. Images illustrate key moments used to manipulate situational urgency.

Button-press behaviour. Participants used the right-hand VR controller to press and hold a trigger whenever they experienced a loss of trust, releasing the button when trust was restored. Button-press time series were aggregated per trial and visualised across urgency and driving agency conditions to characterise moment-to-moment trust dynamics. The button-press mechanism was intentionally designed to capture moments of perceived trust breakdown rather than continuous trust evaluation. It therefore does not capture increases in trust from baseline, but focuses on identifying critical moments of reduced confidence in the driving agent. Prior work suggests that trust in automation evolves dynamically during interaction and may fluctuate in response to confirmations or violations of system expectations (Lee and See, 2004; Kohn et al., 2021). Behavioural measures capturing reliance or intervention behaviours have been widely used as indicators of trust dynamics in human–automation interaction (Kohn et al., 2021). To capture overall trust evaluations, this measure was complemented with post-trial questionnaire ratings.

Trust questionnaire. Trust was assessed using an adapted 7-item Trust in Automation Scale (Jian et al., 2000; Lohaus et al., 2024). Participants completed a baseline questionnaire before the first trial and

Fig. 3. Lab setup used in the study, showing VR headsets, seating arrangement, and driving hardware used during the simulation.

a post-trial questionnaire after each scenario. Items (e.g., "I trusted the automated car/driver") were rated on a 7-point Likert scale. In manual trials, only the two passengers completed the post-trial questionnaire to avoid driver self-assessment bias, ensuring that trust ratings consistently reflected evaluations of an external driving agent.

Measurement symmetry differed across driving conditions. In automated trials, all three participants could indicate trust loss using the button-press mechanism and complete post-trial questionnaires. In manual trials, only passengers provided trust input, as the driver was actively engaged in vehicle control and could not safely perform concurrent self-reporting. This asymmetry reflects a trade-off between ecological validity and measurement symmetry. To account for this, comparisons involving manual driving conditions rely on passenger-only questionnaire responses, ensuring that trust evaluations consistently refer to an external driving agent. Analyses that do not involve manual questionnaire data make use of the full participant sample.

Group conversation. Participants were encouraged to converse naturally throughout each trial. Conversations were recorded and transcribed orthographically using Otter.ai. Transcriptions preserved speech features such as hesitations, false starts, and emphasis to support subsequent qualitative analysis.

3.5. Design

The study employed a 2 × 2 within-subjects design with factors of driving agency (manual vs. automated) and situational urgency (low vs. high). Each triad completed four 10-minute trials (Manual-Low, Manual-High, Automated-Low, Automated-High). Trial order and seat assignments (driver, front passenger, rear passenger) were randomized to minimise potential sequence and learning effects (see Fig. 3). The primary manipulation targeted perceived driving agency rather than driving style; naturalistic variation in manual driving behaviour across triads is acknowledged as a potential confound and is discussed in the Limitations.

3.6. Procedure

Upon arrival, participants provided informed consent and completed a baseline trust questionnaire. After a brief orientation to the equipment, they completed a practice trial to familiarize themselves with the VR environment and button-press mechanism.

During manual trials, one participant controlled the vehicle using a steering wheel and pedals, while the two passengers used VR controllers to report trust loss. Drivers navigated the route using environmental cues (e.g., road signs and barriers) rather than receiving explicit turn-by-turn instructions, in order to preserve more naturalistic navigation behaviour. During automated trials, the vehicle followed a predefined trajectory through the same environment, and all three participants could respond using the button-press mechanism. Participants were instructed to converse naturally during the trials and to press the trigger whenever they experienced a loss of trust in the active driving agent.

Although English proficiency was an inclusion criterion, participants were encouraged to communicate naturally. As all were based in the Netherlands, some triads occasionally switched between English and Dutch. This was permitted only when all three members confirmed Dutch proficiency at the start of the session, ensuring no within-triad comprehension asymmetry. After each trial, participants completed the post-trial trust questionnaire. A scheduled break was provided mid-session, and additional breaks were accommodated on request, during which participants could stand and move around. To mitigate VR-induced discomfort, participants were offered ginger candy and water throughout the session and were consistently reminded that they could pause at any time. The full experimental timeline is shown in Fig. 4.

3.7. Analysis

A multi-method analytical approach was used to examine trust dynamics. Questionnaire data were analysed using repeated-measures ANOVAs to assess the effects of situational urgency and driving agency on trust scores. Paired-samples t-tests were conducted for post-hoc

Table 1
Themes identified in group conversations.

Theme	Description	#C	#G	%G
Navigation uncertainty	Comments about route choice and navigation logic	15	8	66
Automated vehicle behaviour	Shared evaluations of AV performance (speed, lane keeping, manoeuvres)	37	11	91
Environmental factors	References to missing traffic, pedestrians, or ambient realism	8	6	50
Relating to real life	Comparisons to participants' everyday driving experiences	6	6	50
Vehicle feedback	Observations about missing or unclear system cues (e.g., blinkers)	18	8	66
Comparing manual vs. automated	Direct comparisons between perceived control and AV operation	9	7	58

Note. #C = number of conversations; #G = number of groups; %G = percentage of groups referencing theme.

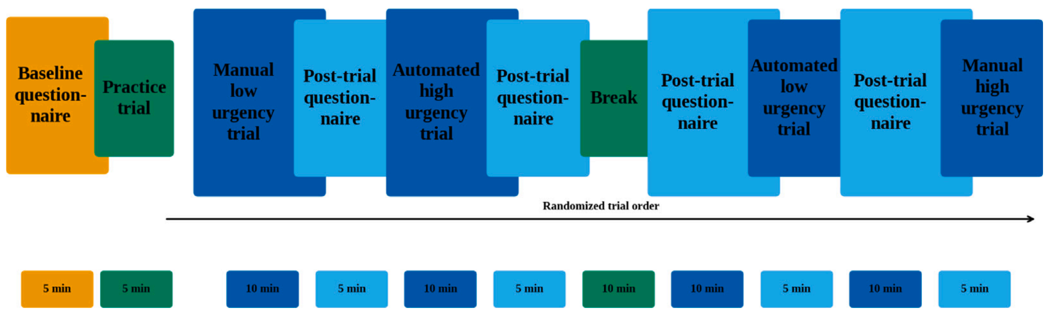

Fig. 4. Timeline of study activities. Participants completed a baseline questionnaire, practice trial, four test trials, and a post-trial questionnaire after each scenario. A break occurred after the second trial.
Note. Trial order and seat assignments were randomized across triads.

comparisons, and an additional ANOVA including the baseline condition was performed to assess trust changes in high-urgency scenarios. Internal consistency of the scales was evaluated using Cronbach's alpha. As data were collected in triads, analyses focused on within-subject effects while treating group context as a shared experimental setting. Additionally, to account for the nested and partially unbalanced data structure, a linear mixed-effects model with random intercepts for participants was conducted as a robustness check.

Button-press behaviour was analysed using repeated-measures ANOVA across the four trial types. Time-locked visualisations illustrated button-press activity in relation to key scenario events. Pearson correlations were calculated between post-trial trust scores and the percentage of trial time in which participants pressed the trust-loss button to assess convergence between subjective and behavioural measures.

Qualitative data were analysed using reflexive thematic analysis following the six-phase framework proposed by Braun and Clarke (2006). Initial codes were generated inductively from the data at both semantic and latent levels and clustered into candidate themes. Themes were reviewed for internal coherence and relevance, then refined and clearly defined. Two researchers independently coded separate halves of the transcripts. Dutch-language transcripts were coded by the Dutch-speaking researcher. To deepen analytical reflexivity, two transcripts from each set were randomly selected for cross-coding between researchers. This procedure was not intended as an inter-rater reliability check, but rather to surface interpretive differences, make analytical assumptions explicit, and refine code definitions through dialogue. Conceptual consensus on theme definitions was reached through iterative discussion rather than statistical agreement thresholds. Final themes were developed through iterative discussion to ensure conceptual coherence and reflexive alignment with the research questions.

In Table 1, prevalence is reported both as the number of conversational instances (#C) in which a theme appeared and as the number and percentage of groups (#G, %G) in which the theme occurred at least once during the trial conversations.

Since analyses were conducted on aggregated trial-level measures, trial order could not be meaningfully included as a predictor and was therefore not included in further analyses.

4. Results

Results are reported across three domains: (1) trust questionnaire scores, (2) button-press behaviour as a real-time indicator of trust loss, and (3) themes identified in group conversations.

4.1. Trust across urgency and driving agency conditions

Internal consistency of the trust questionnaire was high (baseline: $\alpha = 0.93$; post-trial: $\alpha = 0.94$). A paired-samples t-test indicated that trust was significantly higher following high-urgency trials ($M = 4.41$, $SD = 0.97$) compared with low-urgency trials ($M = 4.03$, $SD = 0.97$), $t(35) = 2.39$, $p = .022$, $d = 0.40$.

To examine combined effects of urgency and driving agency, a repeated-measures ANOVA compared four conditions: manual high urgency ($M = 4.67$, $SD = 1.09$), manual low urgency ($M = 4.54$, $SD = 1.23$), automated high urgency ($M = 4.10$, $SD = 1.23$), and automated low urgency ($M = 3.71$, $SD = 1.31$). Participant data for the driver seat were excluded from the analysis of manual trials, as no questionnaire data were available for drivers in these conditions; the analysis was therefore conducted on passenger data only ($N = 24$). Sphericity violations were corrected ($\epsilon = 0.669$). A significant main effect of condition was observed, $F(2.01, 46.17) = 4.80$, $p = .013$, $\eta_p^2 = 0.17$. Bonferroni-adjusted contrasts showed that trust in manual high-urgency trials was significantly higher than in automated low-urgency trials ($\Delta M = 0.96$, $p = .024$). No other pairwise differences reached significance. Fig. 5 illustrates these patterns.

As a robustness check, we accounted for the nested and partially unbalanced structure of the data using a linear mixed-effects model with random intercepts for participants, thereby modelling repeated observations within individuals. Due to model convergence constraints and the limited number of observations per triad, triad-level clustering was not included as an additional random effect. Results showed a

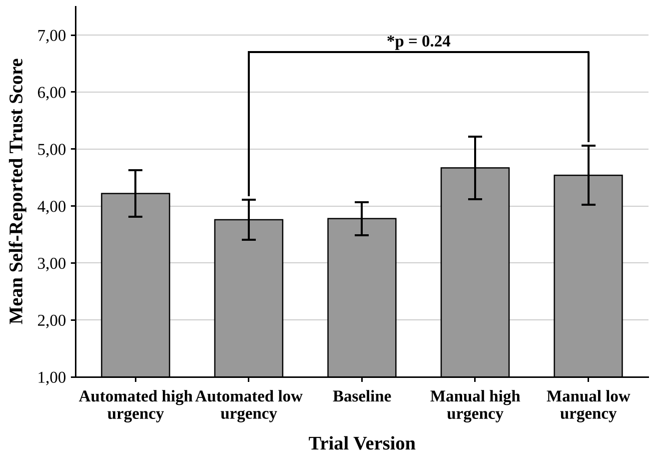

Fig. 5. Trust questionnaire scores across urgency and driving agency conditions. Error bars represent standard error.

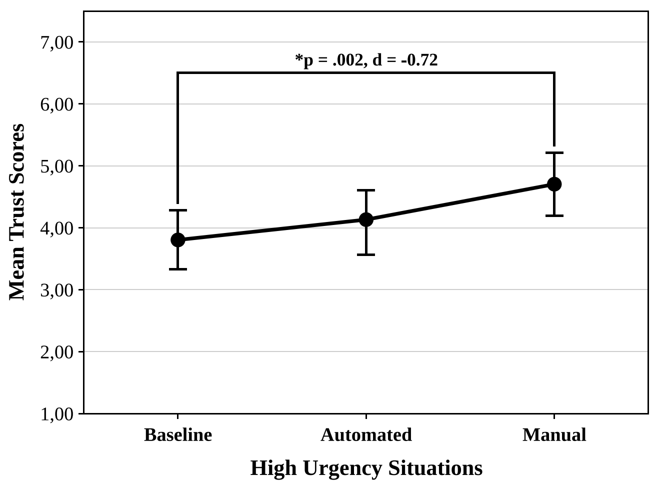

Fig. 6. Estimated marginal means of trust across baseline, automated, and manual conditions in high-urgency scenarios (95% CIs shown).

significant main effect of driving agency, $F(1, 116) = 7.84$, $p = .006$, indicating higher trust in manual compared to automated conditions. No significant main effect of situational urgency, $F(1, 116) = 1.88$, $p = .173$, or interaction between agency and urgency, $F(1, 116) = 0.56$, $p = .457$, was observed. These findings are consistent with the repeated-measures analyses.

To examine trust changes relative to participants' initial expectations, we conducted a follow-up ANOVA including baseline, automated high urgency, and manual high urgency. Results showed a significant effect of condition, $F(2, 46) = 5.82$, $p = .006$, $\eta_p^2 = 0.20$. Trust increased from baseline ($M = 3.77$, $SE = 0.21$) to automated high urgency ($M = 4.05$, $SE = 0.25$), and was highest in manual high urgency ($M = 4.67$, $SE = 0.22$; see Fig. 6). The baseline–manual comparison was significant ($p = .005$, $d = 0.72$); other contrasts were not ($p > .19$).

Additional paired t-tests were conducted to examine trust changes relative to baseline. As only passengers completed the post-trial questionnaire in manual conditions, the baseline-to-manual comparison was restricted to passenger data only ($N = 24$), $t(23) = -3.54$, $p = .002$. The baseline-to-automated comparison included all participants ($N = 36$), $t(35) = -2.01$, $p = .052$, and approached significance only at trend level.

Collapsing across urgency, passenger-reported trust was significantly higher in manual trials ($M = 4.60$, $SD = 1.00$) compared with automated trials ($M = 3.90$, $SD = 1.20$), $t(23) = 2.52$, $p = .019$, $d = 0.51$. Taken together, these results indicate that trust was highest in manual high-urgency scenarios and lowest in automated low-urgency scenarios, suggesting a stronger effect of driving agency than urgency alone.

4.2. Button-press behaviour

Button-press activity captured moment-to-moment trust fluctuations. On average, participants held the trust-loss button for 2.70% of total trial time. Condition means were as follows: automated low urgency ($M = 5.57\%$, $SD = 5.57$), automated high urgency ($M = 2.51\%$, $SD = 2.81$), manual low urgency ($M = 1.55\%$, $SD = 2.46$), and manual high urgency ($M = 1.16\%$, $SD = 1.85$).

A repeated-measures ANOVA revealed a significant effect of condition. Due to sphericity violations ($\chi^2(5) = 57.91$, $p < .001$), Greenhouse–Geisser corrected results are reported: $F(1.47, 48.53) = 15.74$, $p < .001$, $\eta_p^2 = 0.310$.

Bonferroni-corrected comparisons revealed that automated low-urgency trials elicited significantly more button-press behaviour than all other conditions (all $p < .01$). No other contrasts were significant. Fig. 7 shows time-locked peaks aligned with scenario events.

Overall, button-press behaviour was more frequent in automated than manual conditions, with the automated low-urgency scenario generating the highest levels of real-time trust loss.

4.3. Convergence between questionnaire and behavioural measures

To assess convergence between subjective and behavioural indicators of trust, correlations were computed between post-trial questionnaire scores and button-press duration. Only automated trials were included, as they provided button-press data for all participants.

A significant negative correlation emerged, $r(118) = -0.41$, $p < .001$, indicating that lower reported trust was associated with longer button-press durations. Fig. 8 visualises this negative relationship.

These results support button-press behaviour as a valid real-time indicator of trust loss in automated driving contexts.

4.4. Group conversation themes

Thematic analysis of group conversations recorded during the trials revealed several recurrent themes characterising group-level meaning-making processes. Table 1 summarizes the themes, descriptions, and prevalence across conversations.

5. Discussion

This study examined trust in automated mobility as a dynamic and socially embedded process unfolding during shared rides. Across behavioural, subjective, and conversational measures, the findings suggest that trust is not solely an individual evaluation of automation but is shaped by situational context and the perceived source of driving agency. In particular, trust was lowest in automated low-urgency scenarios and highest in manual high-urgency scenarios, indicating that trust calibration depends not only on system behaviour but also on how responsibility for driving is attributed. These findings reinforce the view that trust in automated mobility emerges through an interplay of contextual cues, agency perception, and interpersonal sense-making.

5.1. Urgency, driving agency, and perceived agency

Addressing RQ1, we investigated how trust fluctuated across urgency levels and driving agency. Contrary to initial expectations that high-urgency scenarios would erode trust, trust was higher in high-urgency than in low-urgency trials, particularly when a human was in control. Across measures, participants reported consistently higher trust in manual conditions, suggesting that perceived human agency plays a central role in trust formation.

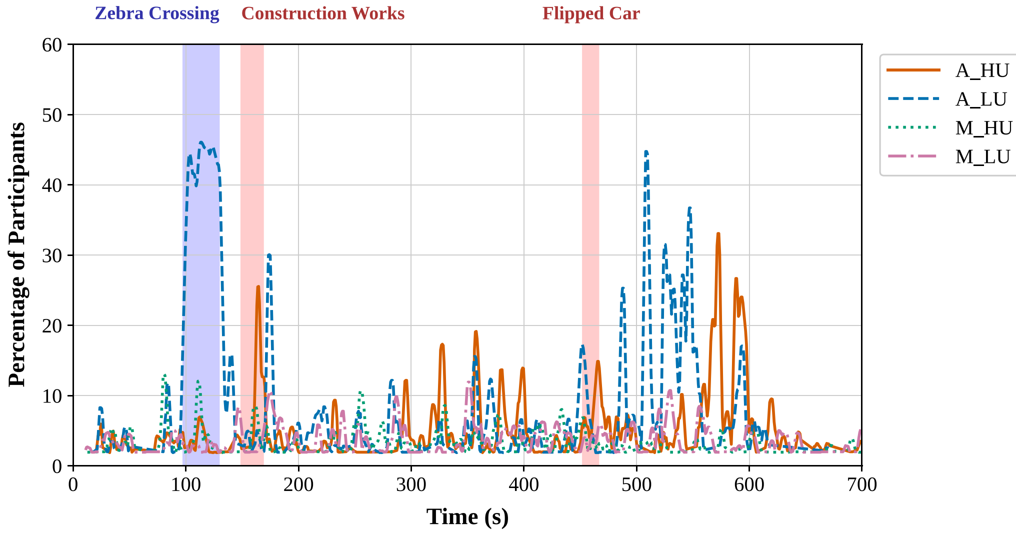

Fig. 7. Percentage of participants holding the trust-loss button over time across trial conditions. Red areas indicate high-urgency events. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

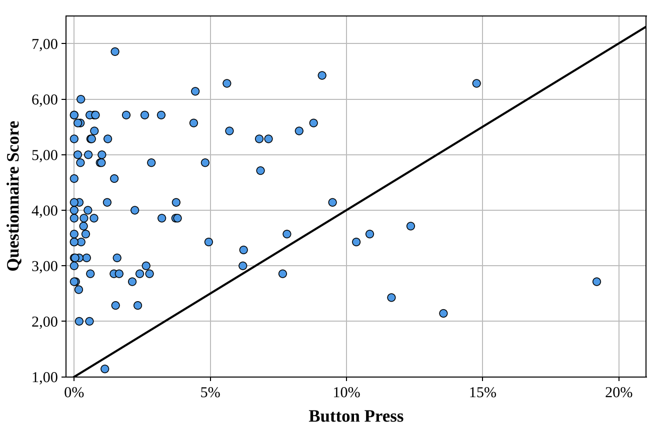

Fig. 8. Correlation between button-press duration and trust scores in automated trials. Each point represents one participant–trial.

Even though only one group member actively controlled the vehicle, the presence of a human driver appeared to reassure the other passengers. This aligns with prior work arguing that trust is shaped not simply by performance, but by how responsibility is distributed between humans and automation (Hoff and Bashir, 2015; Lee and See, 2004; Tan and Zhang, 2025). It is also worth noting that participants' prior indirect exposure to AVs through media and cultural narratives — documented in the Relating to Real Life conversational theme — may have shaped baseline expectations about automated driving capability, contributing to the lower trust observed in automated conditions alongside the perceived agency effect.

Notably, trust in manual high-urgency trials exceeded baseline levels, suggesting that urgency can amplify trust when participants see a human agent managing the situation. By contrast, trust in automated high-urgency trials did not differ significantly from baseline. Urgency alone therefore did not undermine trust, but it also did not substantially deepen it in the absence of perceived human agency. These findings suggest that participants evaluated conditions primarily through a control lens: what mattered was not only what the system did, but who (or what) was perceived to be "in charge".

Group conversations reinforced these perceptions. Participants frequently compared system and manual responses (e.g., "we would have braked sooner"), making control differences highly salient. Low-urgency automated contexts, in particular, seemed to invite scepticism: automation was sometimes perceived as unnecessary or intrusive when the environment did not demand rapid action.

This interpretation is further supported by the qualitative findings. In low-urgency automated trials, participants frequently scrutinised the vehicle's behaviour and questioned its decisions, particularly in relation to routing choices, signalling, and stopping behaviour. Themes such as navigation uncertainty and vehicle feedback illustrate how passengers actively monitored the system and collectively evaluated whether its actions aligned with expectations of competent driving. These conversational reactions suggest that low-demand situations may provide passengers with greater cognitive capacity to observe and critique automated behaviour, leading to increased scepticism when automation appears unnecessary or insufficiently transparent. Taken together, the behavioural signals, questionnaire ratings, and conversational data converge to indicate that trust breakdowns were most likely to occur when automation was perceived as unjustified in otherwise low-risk situations.

These findings also resonate with broader discussions of responsibility attribution in human–automation interaction. Trust appeared to be closely linked to perceptions of who was "in charge" of the driving task, with participants expressing greater comfort when responsibility was clearly human and more scepticism when control was delegated to automation without transparent feedback. This aligns with work on responsibility allocation and moral dilemma framing in automated driving, which shows that perceptions of agency and accountability shape how people evaluate automated systems and their decisions (Yoo et al., 2021). In this sense, trust in automated mobility may be inseparable from how responsibility is perceived, negotiated, and socially interpreted during shared rides. Whereas prior research has primarily examined responsibility attribution through abstract ethical scenarios or survey-based dilemmas, the present study demonstrates how perceptions of control and accountability are negotiated dynamically through real-time interaction and group discussion during shared automated rides.

These findings can additionally be interpreted through the lens of passengers' sense of individual autonomy. When a human driver was present, passengers retained indirect means of influencing the journey: they could voice concern, prompt a response through conversation, or exert social authority over the situation. Manual conditions may therefore have preserved felt autonomy even for non-driving occupants. Automated conditions removed this channel, leaving passengers without a clear mechanism to assert preference or intervene. The resulting reduction in felt autonomy may have compounded distrust — particularly in low-urgency scenarios, where passengers had the cognitive capacity to notice and reflect on their limited influence, and where the system's actions appeared unnecessary or opaque. This suggests that trust in shared AVs may be partly a function of perceived individual autonomy: not simply the competence or agency of the driving entity, but the passenger's own sense of having a voice in the journey. Given that limited prior work has directly examined this dimension in AV trust research, the present findings offer an initial empirical basis for motivating autonomy as a meaningful construct in this domain.

5.2. Button presses as real-time indicators of trust

Addressing RQ2, we evaluated whether button-press behaviour functioned as a real-time indicator of trust loss. The highest levels of button-press activity occurred in automated low-urgency trials, where participants likely had greater cognitive capacity to monitor the vehicle's behaviour and question its decisions. This condition also showed the greatest variability in button-press duration, suggesting that some participants reacted strongly while others did not press at all. Such variability is consistent with models of cognition-based trust (Hoff and Bashir, 2015), which emphasize individual differences in how users evaluate automation under low-stress conditions.

Button-press patterns converged with questionnaire data, particularly in automated trials, where a significant negative correlation between button-press duration and post-trial trust scores indicated that longer button holding was associated with lower reported trust. This supports the interpretation of button presses as a meaningful behavioural index of trust loss, rather than a noisy or purely exploratory signal.

However, the qualitative data suggest that button presses were embedded in social practices and may have been shaped by interpersonal dynamics. Participants occasionally discussed whether or when to press, effectively treating the mechanism as a shared monitoring tool rather than a strictly individual signal. This raises the possibility of social influence effects, where one passenger's verbalized doubt could prompt others to signal trust loss. In addition, participants may have interpreted the instruction differently, responding not only to loss of trust but also to surprise, novelty, or momentary uncertainty in the vehicle's behaviour. In this sense, button-press behaviour may capture both individual trust fluctuations and socially coordinated trust calibration processes. While this introduces potential demand characteristics, it also reflects the inherently social nature of trust in shared mobility contexts, where perceptions are co-regulated through conversation and mutual monitoring.

The button-press mechanism provides an important contribution by capturing micro-temporal expressions of trust breakdown that are difficult to obtain through retrospective measures. While such real-time inputs may be influenced by social dynamics (e.g., participants reacting to each other's uncertainty), this should not be viewed solely as a limitation. In this sense, button-press behaviour may capture not only individual trust loss but also collective trust calibration processes that unfold during shared automated travel.

5.3. Social dynamics and shared trust formation

To address RQ3, we examined how trust in AV performance was co-constructed through group interaction. Thematic analysis identified six recurrent categories of talk that captured how passengers collaboratively made sense of the ride: navigation uncertainty, automated vehicle behaviour, environmental factors, relating experiences to real life, vehicle feedback, and comparisons between automated and manual modes. These findings illustrate that passengers did not form trust judgements in isolation. Instead, they jointly interpreted events, negotiated meaning, and aligned their responses, creating a shared trust trajectory throughout each session.

Quotes follow the convention: G = group number, T = trial type (TA = automated, TM = manual), and seat number (0 = driver, 1 = front passenger, 2 = rear passenger). Time stamps (mm:ss) indicate when the statement occurred in the trial. Representative excerpts illustrating the six themes are summarized in Fig. 9.

Navigation uncertainty. Across eight groups (66%), passengers expressed uncertainty about the route, destination, or directional choices. These conversations frequently emerged when the vehicle appeared to circle the same area or when the next turn was ambiguous. For example, one participant remarked, *"Like we've circled around, there's no destination"* (G1, TA, 0, 10:37), followed by another noting, *"There's no navigation"* (G1, TA, 1, 10:40). Similar confusion appeared in manual trials, as when a driver asked, *"Where am I supposed to go here, left or right?"* (G2, TM, 0, 20:20). These exchanges exemplify collective sense-making under uncertainty and highlight how ambiguity can spread socially, amplifying group-level anxiety in line with collective risk perception (Momen et al., 2023; Loersch et al., 2008). From an affective trust perspective (Lee and See, 2004), ambiguous routing heightened emotional unease, undermining confidence in the system.

Automated vehicle behaviour. Nearly all groups (91%) scrutinised the AV's driving behaviour, commenting on lane changes, speed variation, stopping logic, and the absence of signals. For example, in one triad participants exclaimed, *"Oh, we are going straight. Wait, what the heck?"* (G1, TA, 2, 11:10), followed by *"Still no turning signal"* (G1, TA, 0, 11:19). Concerns about unsafe or unclear stopping also surfaced: *"It's not the best place to stop"* (G5, TA, 2, 49:27). However, not all evaluations were negative; some acknowledged smoothness, such as *"It's driving very calm"* (G2, TA, 0, 3:30). These judgements reflect analytical trust processing (Lee and See, 2004), as participants compared the system's actions to implicit rules of "good driving". Behavioural inconsistencies, however, invited group-level critique, consistent with findings that opaque or non-normative automation behaviour erodes trust (Domeyer et al., 2018).

Environmental factors. Half the groups (50%) commented on environmental realism, especially the absence of other vehicles or pedestrians. This absence was sometimes experienced as eerie or unrealistic. For instance, participants in G4 questioned, *"Why are there no other cars?"* (G4, TA, 2, 2:27), followed by *"It's creepy"* (G4, TA, 0, 2:30). Others discussed how emptiness reduced perceived risk: *"If there's no other cars around, I'm not really afraid"* (G5, TA, 2, 7:24). These reactions underscore the importance of ecological validity in shaping trust calibration (Tran et al., 2021). Simplified environments limited participants' ability to assess decision-making complexity, thereby reducing trust or producing discomfort.

Relating to real life. Six groups (50%) contextualized the simulation by referencing real-world experiences, cultural driving norms, or media portrayals of AVs. For instance, one participant asked, *"If you have an automated car, is it legal to drink and drive?"* (G2, TA, 2, 11:02), while another envisioned future interiors: *"Self-driving car could be like four chairs facing each other"* (G3, TA, 0, 56:28). Others mentioned videos of driverless taxis: *"I have seen some on TikTok ... there's literally no person inside"* (G4, TA, 03:27). These analogical reasoning processes (Lee and See, 2004) demonstrate how participants used prior experiences and cultural narratives to interpret system behaviour, highlighting how trust evolves through both direct interaction and indirect knowledge (Guo et al., 2023).

Vehicle feedback. Eight groups (66%) emphasized the centrality of feedback, particularly signalling and anticipatory cues, for maintaining trust. Missing blinkers were repeatedly reported: *"Still no turn signals"* (G1, TA, 2, 8:39), followed by *"Seems like a major oversight"* (G1, TA, 2, 8:57). Unexplained stops further heightened uncertainty: *"What's going on? Why did we stop?"* (G2, TA, 2, 4:07). These concerns align with literature on transparency and predictability in automated systems

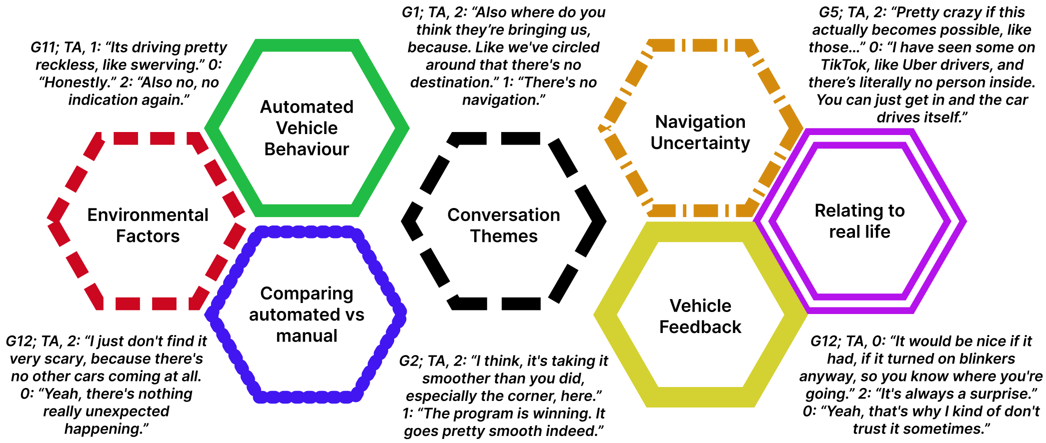

Fig. 9. Representative participant quotes illustrating key aspects of trust co-construction across the six qualitative themes.
Note. Quotes follow the coding scheme: G = group number, *T* = trial type, and seat number (0 = driver, 1 = front passenger, 2 = rear passenger).

(Atakishiyev et al., 2024). The group-level nature of these reactions suggests that insufficient feedback undermines trust collectively—when one participant voiced doubt, others quickly reinforced or expanded the critique.

Comparing automated vs. manual. Seven groups (58%) explicitly compared manual and automated driving. Manual driving was sometimes perceived as more competent or predictable: *"Much better"* (G1, TM, 1, 26:21). Yet, participants also noted advantages of automation: *"I think it's taking it smoother than you did"* (G1, TA, 2, 22:00). Comparisons often became humorous or evaluative exchanges, such as noting the human driver's lapses: *"You weren't looking at all the blind spots"* (G2, TM, 2, 33:05). These dialogues illustrate analogical trust processes (Lee and See, 2004), showing how participants used human driving as a yardstick to benchmark the AV, resulting in nuanced, negotiated trust calibration within the group.

Overall, the qualitative findings position trust in AVs as a socially embedded process. Group members used conversation to interpret behaviour, contextualize risk, and align responses, with individual perceptions and system cues feeding into shared meaning-making. Social dynamics were thus present throughout trust formation during automated and manual rides, providing a collective context within which individual trust responses were expressed and interpreted. It is worth distinguishing this sense-making process from social influence: whereas sense-making refers to the joint, emergent construction of meaning from ambiguous events, social influence involves directional transmission of attitudes or reactions from one group member to another. The present data document primarily the former, though the button-press findings suggest that influence effects may also operate within these interactions.

Taken together, the three measurement modalities provide complementary insight into the dynamics of trust formation during shared rides. Real-time button presses captured immediate reactions to perceived anomalies in vehicle behaviour, while group conversations revealed how passengers interpreted and negotiated these moments collectively. Post-trial questionnaire ratings then reflected participants' retrospective evaluations of the driving agent. Viewed together, these measures illustrate how trust unfolds through a sequence of perception, social interpretation, and reflective assessment, highlighting the value of combining behavioural, subjective, and conversational data to study trust as a temporally evolving and socially embedded process.

6. Implications for design

The findings indicate that trust in AVs is context-sensitive, socially constructed, and partially accessible through behavioural signals. These insights have direct relevance for the design and deployment of SAVs in real-world transport contexts.

The qualitative themes summarized in Table 1—such as navigation uncertainty, missing vehicle feedback cues, and comparisons between automated and manual driving—provide concrete insight into the situations in which passengers questioned system behaviour. The following design implications translate these conversational patterns into actionable recommendations for SAV interfaces.

6.1. Support real-time trust expression

The button-press mechanism functioned as a research instrument in this study, but its consistent and meaningful use by participants motivates a broader design principle: future AV systems should incorporate unobtrusive channels through which passengers can signal reduced confidence or discomfort in real time. Rather than relying solely on automated sensing, such inputs could allow the system to respond adaptively to passenger state — for example by modulating speed, providing explanatory feedback, or flagging collective unease to a remote operator. Time-locked peaks around critical moments demonstrate the potential of such mechanisms for monitoring passenger state, and future implementations could include haptic, touch-based, or gesture-based inputs that allow passengers to express discomfort without disrupting ongoing interaction. These inputs could serve as both diagnostic tools for designers and triggers for adaptive system responses.

6.2. Design for social meaning-making

Themes such as navigation uncertainty and vehicle feedback (Table 1) showed that passengers frequently relied on each other to interpret AV behaviour, particularly in ambiguous moments. Interfaces should therefore support shared understanding, not just individual comprehension. Ambient feedback such as clear signalling, route visualisations, and brief verbal or visual explanations of upcoming manoeuvres can provide common ground for group discussion and reduce unnecessary speculation. Designing displays and feedback visible to all occupants, rather than only the driver's position, may be especially important in SAVs.

6.3. Enhance contextual responsiveness

Consistent with the themes of navigation uncertainty and automated vehicle behaviour (Table 1), trust was consistently lowest in automated low-urgency scenarios, suggesting that perceived redundancy, unexplained pauses, or conservative manoeuvres may be interpreted as poor performance when the environment appears benign. AVs may need to adapt their communication strategies to context: in low-event-density situations, where the vehicle's actions may seem opaque or unnecessary, additional explanatory feedback (e.g., "Slowing due to speed limit change ahead") could help maintain trust. In higher-urgency situations, timely and appropriately calibrated feedback may support the sense that the system is attentive and in control.

6.4. Account for group settings and multimodal input

Given that shared rides involve multiple occupants co-constructing trust, future SAVs should be able to detect and respond to group-level dynamics. Combining multimodal indicators, such as aggregated button presses, gaze patterns, or vocal cues, may enable the system to infer when a group is collectively uneasy or divided. Adaptive transparency strategies (e.g., offering more detailed explanations when uncertainty appears high) could help stabilize trust in these moments. Designing features that acknowledge the group, rather than addressing only an individual, may further support shared trust.

6.5. Strengthen ecological validity in design and testing

The environmental factors theme (Table 1), where participants frequently remarked on the absence of other traffic and pedestrians, highlights the importance of ecological validity in shaping trust. For real-world deployment, AV design should align behavioural and interface cues with passengers' expectations of a rich, populated environment. Even in simulation and testing phases, including more realistic traffic actors, ambient sound, and standard safety cues (e.g., seatbelt indicators, audible alerts) can support more natural calibration of trust and make findings more transferable to real-world settings.

7. Limitations and future work

Several limitations should be considered when interpreting the findings of this study.

The coordinates and movements of the simulated vehicle were not logged. As a result, participants' real-time trust-loss signals (e.g., button presses) could not be aligned precisely with specific environmental events. This limits the ability to conduct event-based analyses of trust dynamics. Future research should incorporate synchronized trajectory logging and time-stamped event markers to link behavioural responses directly to road situations, such as lane changes, obstacles, or unexpected stops.

Another limitation concerns the absence of an explicit manipulation check for situational urgency. Although the experimental design introduced clear event-based contrasts (routine navigation vs. sudden hazards), perceived urgency was not directly measured through self-report ratings (e.g., perceived time pressure or stress). Instead, the effectiveness of the manipulation was indirectly supported by behavioural and conversational patterns observed across conditions. However, without a direct manipulation check, conclusions regarding perceived urgency should be interpreted with caution. Future work should include explicit self-report measures to strengthen construct validity and allow more precise assessment of perceived urgency.

The measurement of real-time trust was asymmetric across conditions. In manual trials, only passengers could press the trust-loss button, whereas drivers could not signal reduced trust while actively operating the vehicle. This limits comparability across driving agency conditions and may underestimate moments of low trust in manual driving. Subsequent studies should employ more balanced multimodal sensing strategies (e.g., haptics, physiological indicators, or verbal markers) that capture trust-related responses from all occupants, including drivers. A further constraint is that the button-press mechanism captures trust loss only; trust gain during a trial is inferred indirectly from button release events and from the proportion of trial time spent not pressing, rather than from a direct positive input. Future work could explore bidirectional real-time trust signalling to provide more complete coverage of within-trial trust dynamics. Additionally, the linear mixed-effects model used as a robustness check included random intercepts for participants but could not model triad-level random effects, as the study comprised only 12 groups — insufficient to reliably estimate between-group variance; group-level variability is therefore not fully partitioned in the model.

A further limitation concerns potential confounding from driving style variation. Manual driving was not experimentally matched to the automated condition in terms of speed, braking, or lateral behaviour, as the design prioritised naturalistic driving over stylistic control. Differences in trust between conditions may therefore partly reflect individual variation in manual driving behaviour rather than driving agency alone. Future studies should consider standardising key driving parameters across conditions, or systematically varying driving style, to isolate its contribution to trust perceptions.

The simulation also lacked certain elements of ecological realism. The absence of other traffic participants, missing safety features (such as visible seatbelts), and occasional interface constraints may have influenced how participants calibrated their trust. Some aspects of vehicle behaviour, such as limited signalling or stopping decisions, may also have appeared unrealistic. While these were not designed as explicit experimental manipulations, they formed part of the environment that participants interpreted and discussed during the trials. These factors could have attenuated perceived risk or made the environment feel artificial or sparsely populated. This may have been particularly relevant in low-urgency trials, where the absence of surrounding traffic could make automated behaviour appear unnecessary and invite closer scrutiny from passengers. Increasing environmental fidelity—by adding surrounding traffic, pedestrians, ambient sound, and standard in-vehicle safety cues, and ensuring consistent exposure to urgency events—would improve generalizability and support more robust comparisons across conditions.

Another limitation concerns the participant sample. Participants were recruited primarily from university student populations through research credit systems, including undergraduate and graduate students from technical and social science programs. Student samples may differ from the broader population in their familiarity with emerging technologies and openness to automated systems, which could influence trust judgements. While student participants are commonly used in experimental VR research, future studies should examine whether similar trust dynamics emerge in more diverse demographic groups.

A small number of participants experienced VR-induced sickness severe enough to discontinue the study early. These participants received one research credit as partial compensation, and their incomplete data were excluded from analysis. While mitigation strategies were employed throughout — including offering ginger candy and water, providing flexible breaks on request, and allowing participants to stand and move between trials — VR sickness remains an inherent limitation of immersive simulation research and may have introduced a degree of self-selection bias, as participants more susceptible to motion sickness were unable to complete all trials.

Prior familiarity between group members was not systematically recorded, representing a limitation for interpreting the qualitative findings in particular. Groups in which participants already knew each other may have exhibited different conversational dynamics, levels of candour, and willingness to express doubt compared with groups of strangers, potentially influencing how trust was socially expressed and interpreted. Future work should systematically record and control for prior familiarity to examine its role in shaping group trust dynamics.

Finally, conversational engagement varied substantially between groups. Since participants were based in the Netherlands, some groups also occasionally switched between English and Dutch during the conversations. Although this allowed participants to communicate naturally, language switching may have influenced conversational dynamics by shaping how uncertainty, humour, or critique of the vehicle was expressed. Some triads engaged in rich, continuous discussion, whereas others remained relatively quiet. These differences likely reflected variation in group cohesion, comfort in VR, and individual communication styles, and they may have influenced how trust was socially constructed. Furthermore, conversational initiation patterns and turn-taking sequences were not coded, meaning that leader–follower dynamics — where one group member's expressed evaluation is confirmed by others — cannot be ruled out as a contributing factor to the group-level patterns observed. Future work should systematically investigate how group composition and interaction quality shape shared trust formation, for example through linguistic style matching, conversation metrics, and multimodal behavioural cues, and employ sequential or discourse analysis methods to allow a more precise characterisation of whether observed convergence reflects joint sense-making or conformity to an initially dominant view. Additionally, within-triad trust profiles — for example, whether heterogeneous configurations of high- and low-trust members produced distinct conversational dynamics such as reassurance or resistance — were not examined; future work mapping individual button-press patterns onto conversation transcripts could yield important insights into trust contagion and influence processes within shared rides.

Taken together, these limitations underscore that trust in AVs is shaped not only by situational and technological factors, but also by methodological design choices and group dynamics. Future studies should build on these insights using higher-fidelity environments, balanced measurement strategies, and more systematic manipulation of group characteristics to better understand how trust develops in shared automated mobility.

8. Conclusion

This study examined how trust in AVs develops during shared rides by investigating the combined effects of driving agency, situational urgency, and group interaction. Across measures, trust was consistently higher in manual than automated conditions and lowest in automated low-urgency scenarios, where system behaviour appeared less necessary or less transparent. High-urgency situations increased trust primarily when a human was perceived to be in control, underscoring the importance of perceived agency and responsibility in trust formation.

Real-time button-press behaviour provided a sensitive indicator of moment-to-moment trust loss, particularly in automated trials. Its convergence with questionnaire scores demonstrates the value of combining behavioural and self-report measures to capture dynamic trust fluctuations in automated mobility contexts. In addition, qualitative findings showed that trust judgements were shaped through group-level sense-making, with passengers jointly interpreting vehicle behaviour, ambiguity, and perceived control.

Overall, the findings position trust in automated mobility as a socially embedded phenomenon shaped by perceived agency, contextual dynamics, and interpersonal interaction. In shared ride settings, passengers continuously interpreted system behaviour within a social context, negotiating meaning, validating impressions, and regulating uncertainty, suggesting that purely individual models of trust in automation may not fully capture how trust unfolds in shared rides. By integrating behavioural, subjective, and conversational evidence, this study demonstrates the value of multi-method approaches for understanding trust beyond individual-level models. As automated mobility evolves towards shared deployment, future systems must account for collective trust formation processes and design for transparency, shared understanding, and adaptive responsiveness in group settings.

Beyond interface design, these findings carry implications for transport policy and the governance of shared automated mobility. As SAV services move from pilot programs towards public deployment, including robotaxi schemes and autonomous ride-sharing platforms, policymakers must consider not only technical safety standards but also conditions for collective trust formation. Regulatory frameworks that require AVs to provide transparent, group-accessible feedback and that evaluate passenger trust under shared rather than individual conditions would better reflect the social reality of how trust develops in these contexts. The present findings support the case for including social and behavioural evidence alongside engineering metrics in SAV certification and deployment assessments.

CRediT authorship contribution statement

Ugnė Aleksandra Morkutė: Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Pavlo Bazilinskyy:** Writing – review & editing, Supervision, Software, Resources, Methodology, Conceptualization. **Jom Dieben:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Francesco Walker:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Ethics statement

The study was approved by the Psychology Research Ethics Committee of [anonymous university]. All participants provided written informed consent prior to participation. The study was conducted in accordance with the ethical standards of the institutional research committee and the Declaration of Helsinki.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Claude (Anthropic) to assist with manuscript reformatting, citation style conversion, and language editing. After using this tool, the authors reviewed and edited all content and take full responsibility for the content of the published article.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank all participants for their time and valuable contributions to this study. We are also grateful to the friends and colleagues who supported the pilot testing phase. The authors extend their appreciation to the members of the lab for their technical assistance and ongoing support throughout the project.

Data availability

All data and materials supporting this study are available on the Open Science Framework (OSF): https://osf.io/7d4mw. The repository includes the baseline trust questionnaire, the manual and automated post-trial questionnaires, the scoring key used to compute composite trust scores, the timeline of critical events for the automated conditions, all processed data used to generate the analyses and figures, and the full analysis code. All materials are provided exactly as used in the study.

References

Adnan, N., 2024. Exploring the future: a meta-analysis of autonomous vehicle adoption and its impact on urban life and the healthcare sector. Transp. Res. Interdiscip. Perspect. 26, 101110. http://dx.doi.org/10.1016/j.trip.2024.101110.

Atakishiyev, S., Salameh, M., Goebel, R., 2024. Incorporating explanations into human–machine interfaces for trust and situation awareness in autonomous vehicles. In: Proceedings of the 2024 IEEE Intelligent Vehicles Symposium. pp. 2948–2955. http://dx.doi.org/10.1109/IV55156.2024.10588812.

Ayoub, J., Avetisian, L., Yang, X.J., Zhou, F., 2023. Real-time trust prediction in conditionally automated driving using physiological measures. IEEE Trans. Intell. Transp. Syst. 24 (12), 14642–14650. http://dx.doi.org/10.1109/TITS.2023.3295783.

Baccari, S., Hadded, M., Ghazzai, H., Touati, H., Elhadef, M., 2024. Anomaly detection in connected and autonomous vehicles: a survey, analysis, and research challenges. IEEE Access 12, 19250–19276. http://dx.doi.org/10.1109/ACCESS.2024.3361829.

Bagloee, S.A., Tavana, M., Asadi, M., Oliver, T., 2016. Autonomous vehicles: challenges, opportunities, and future implications for transportation policies. J. Mod. Transp. 24 (4), 284–303. http://dx.doi.org/10.1007/s40534-016-0117-3.

Bazilinskyy, P., Kooijman, L., Dodou, D., De Winter, J.C.F., 2020. Coupled simulator for research on the interaction between pedestrians and (automated) vehicles. In: Proceedings of the Driving Simulation Conference. Antibes, France.

Bazilinskyy, P., Kyriakidis, M., de Winter, J., 2015. An international crowdsourcing study into people's statements on fully automated driving. Procedia Manuf. 3, 2534–2542. http://dx.doi.org/10.1016/j.promfg.2015.07.540.

Biondi, F., Alvarez, I., Jeong, K., 2019. Human–vehicle cooperation in automated driving: a multidisciplinary review and appraisal. Int. J. Hum.–Comput. Interact. 35 (11), 932–946. http://dx.doi.org/10.1080/10447318.2018.1561792.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3 (2), 77–101. http://dx.doi.org/10.1191/1478088706qp063oa.

Choi, J.K., Ji, Y.G., 2015. Investigating the importance of trust on adopting an autonomous vehicle. Int. J. Hum.–Comput. Interact. 31 (10), 692–702. http://dx.doi.org/10.1080/10447318.2015.1070549.

Cooke, N.J., 2015. Team cognition as interaction. Curr. Dir. Psychol. Sci. 24 (6), 415–419. http://dx.doi.org/10.1177/0963721415602474.

De Cet, G., Orsini, F., Meneguzzer, C., Gastaldi, M., Saljoqi, M., Rossi, R., 2024. Do we trust automated vehicles? A driving simulator study. Transp. Res. Procedia 78, 174–181. http://dx.doi.org/10.1016/j.trpro.2024.02.023.

Dikmen, M., Burns, C., 2017. Trust in autonomous vehicles: the case of Tesla Autopilot and Summon. In: IEEE International Conference on Systems, Man and Cybernetics. SMC, pp. 1093–1098. http://dx.doi.org/10.1109/SMC.2017.8122757.

Dolins, S., Strömberg, H., Wong, Y.Z., Karlsson, M., 2021. Sharing anxiety is in the driver's seat: analyzing user acceptance of dynamic ridepooling and its implications for shared autonomous mobility. Sustainability 13 (14), 7828. http://dx.doi.org/10.3390/su13147828.

Domeyer, J., Venkatraman, V., Price, M., Lee, J.D., 2018. Characterizing driver trust in vehicle control algorithm parameters. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 62 (1), 1821–1825. http://dx.doi.org/10.1177/1541931218621413.

Fagnant, D.J., Kockelman, K., 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transp. Res. A 77, 167–181. http://dx.doi.org/10.1016/j.tra.2015.04.003.

Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39 (2), 175–191. http://dx.doi.org/10.3758/BF03193146.

Felbel, K., Dettmann, A., Bullinger, A.C., 2022. Analysis of eye gaze given different automated driving styles in an urban environment. In: Proceedings of the 2022 IEEE 9th International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications. http://dx.doi.org/10.1109/CIVEMSA53371.2022.9853647.

Fu, E., Hyde, D., Sibi, S., Johns, M., Fischer, K., Sirkin, D., 2020. Assessing the effects of failure alerts on transitions of control from autonomous driving systems. In: Proceedings of the 2020 IEEE Intelligent Vehicles Symposium. pp. 1956–1963. http://dx.doi.org/10.1109/IV47402.2020.9304627.

Gold, C., Körber, M., Hohenberger, C., Lechner, D., Bengler, K., 2015. Trust in automation – before and after take-over scenarios in a highly automated vehicle. Procedia Manuf. 3, 3025–3032. http://dx.doi.org/10.1016/j.promfg.2015.07.847.

Guo, Y., Yang, X.J., Shi, C., 2023. TIP: a trust inference and propagation model in multi-human multi-robot teams. In: Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. pp. 639–643. http://dx.doi.org/10.1145/3568294.3580164.

Hergeth, S., Lorenz, L., Vilimek, R., Krems, J.F., 2016. Keep your scanners peeled: gaze behavior as a measure of automation trust during highly automated driving. Hum. Factors 58 (3), 509–519. http://dx.doi.org/10.1177/0018720815625744.

Hoff, K.A., Bashir, M., 2015. Trust in automation: integrating empirical evidence on factors that influence trust. Hum. Factors 57 (3), 407–434. http://dx.doi.org/10.1177/0018720814547570.

Hossain, M., Abdel-Aty, M., Quddus, M.A., Muromachi, Y., Sadeek, S.N., 2019. Real-time crash prediction models: state-of-the-art, design pathways and ubiquitous requirements. Accid. Anal. Prev. 124, 66–84. http://dx.doi.org/10.1016/j.aap.2018.12.022.

Jian, J.-Y., Bisantz, A.M., Drury, C.G., 2000. Foundations for an empirically determined scale of trust in automated systems. Int. J. Cogn. Ergon. 4 (1), 53–71. http://dx.doi.org/10.1207/S15327566IJCE0401_04.

Ju, U.-L., Chuang, L.L., Wallraven, C., 2022. Acoustic cues increase situational awareness in accident situations: a VR car-driving study. IEEE Trans. Intell. Transp. Syst. 23 (4), 3281–3291. http://dx.doi.org/10.1109/TITS.2020.3035374.

Kaufman, R.A., Lee, E., Bedmutha, M.S., Kirsh, D., Weibel, N., 2025. Predicting trust in autonomous vehicles: modeling psychosocial traits, risk-benefit attitudes, and driving factors with machine learning. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. pp. 1–24. http://dx.doi.org/10.1145/3706598.3713188.

Kohn, S.C., de Visser, E.J., Wiese, E., Lee, Y.-C., Shaw, T.H., 2021. Measurement of trust in automation: a narrative review and reference guide. Front. Psychol. 12, http://dx.doi.org/10.3389/fpsyg.2021.604977.

Kraus, J., Scholz, D., Stiegemeier, D., Baumann, M., 2020. The more you know: trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency. Hum. Factors 62 (5), 718–736. http://dx.doi.org/10.1177/0018720819853686.

Lee, J.D., See, K.A., 2004. Trust in automation: designing for appropriate reliance. Hum. Factors 46 (1), 50–80. http://dx.doi.org/10.1518/hfes.46.1.50_30392.

Loersch, C., Aarts, H., Payne, B.K., Jefferis, V.E., 2008. The influence of social groups on goal contagion. J. Exp. Soc. Psychol. 44 (6), 1555–1558. http://dx.doi.org/10.1016/j.jesp.2008.07.009.

Lohaus, L., Woide, M., Damm, N., Demiral, Z., Friedrich, H., Petáková, A., Walker, F., 2024. Automated or human: which driver wins the race for the passengers' trust? Examining passenger trust in human-driven and automated vehicles following a dangerous situation. Comput. Hum. Behav. 161, 108387. http://dx.doi.org/10.1016/j.chb.2024.108387.

Lu, Y., Sarter, N., 2019. Eye tracking: a process-oriented method for inferring trust in automation as a function of priming and system reliability. IEEE Trans. Hum.-Mach. Syst. 49 (6), 560–568. http://dx.doi.org/10.1109/THMS.2019.2930980.

Meteier, Q., Capallera, M., de Salis, E., Angelini, L., Carrino, S., Widmer, M., Abou Khaled, O., Mugellini, E., Sonderegger, A., 2023. A dataset on the physiological state and behavior of drivers in conditionally automated driving. Data Brief 47, 109027. http://dx.doi.org/10.1016/j.dib.2023.109027.

Miskolczi, M., Kökény, L., Jászberényi, M., 2025. Rethinking the road ahead – generation Z's perspectives on AI-based mobility services. Transp. Res. Interdiscip. Perspect. 31, 101475. http://dx.doi.org/10.1016/j.trip.2025.101475.

Momen, A., de Visser, E.J., Fraune, M.R., Madison, A., Rueben, M., Cooley, K., Tossell, C.C., 2023. Group trust dynamics during a risky driving experience in a Tesla Model X. Front. Psychol. 14, http://dx.doi.org/10.3389/fpsyg.2023.1129369.

Muir, B.M., 1994. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics 37 (11), 1905–1922. http://dx.doi.org/10.1080/00140139408964957.

National Highway Traffic Safety Administration, 2008. National Motor Vehicle Crash Causation Survey: Report to Congress. Technical Report DOT HS 811 059, U.S. Department of Transportation, Washington, DC.

Nordhoff, S., Stapel, J., He, X., Gentner, A., Happee, R., 2021. Perceived safety and trust in SAE Level 2 partially automated cars: results from an online questionnaire. PLoS ONE 16 (12), e0260953. http://dx.doi.org/10.1371/journal.pone.0260953.

Perello-March, J.R., Burns, C.G., Woodman, R., Elliott, M.T., Birrell, S.A., 2023. Using fNIRS to verify trust in highly automated driving. IEEE Trans. Intell. Transp. Syst. 24 (1), 739–752. http://dx.doi.org/10.1109/TITS.2022.3211089.

Pettersson, I., Karlsson, I.C.M., 2015. Setting the stage for autonomous cars: a pilot study of future automated driving experiences. IET Intell. Transp. Syst. 9 (7), 694–701. http://dx.doi.org/10.1049/iet-its.2014.0168.

Raats, K., Fors, V., Pink, S., 2020. Trusting autonomous vehicles: an interdisciplinary approach. Transp. Res. Interdiscip. Perspect. 7, 100201. http://dx.doi.org/10.1016/j.trip.2020.100201.

Rojas, E., Hsu, D., Huang, J., Li, M., 2025. Interpersonal influence matters: trust contagion and repair in human-human-AI team. Comput. Hum. Behav.: Artif. Hum. 5, 100194. http://dx.doi.org/10.1016/j.chbah.2025.100194.

Ryan, R.M., Deci, E.L., 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. 55 (1), 68–78. http://dx.doi.org/10.1037/0003-066X.55.1.68.

Shepardson, D., 2024. GM Self-Driving Unit Cruise Admits to Submitting False Report, Will Pay $500,000 Fine. Reuters, https://tinyurl.com/reuters-cruise-2024 (November 2024).

Strauch, C., Mühl, K., Patro, K., Grabmaier, C., Reithinger, S., Baumann, M., Huckauf, A., 2019. Real autonomous driving from a passenger's perspective: two experimental investigations using gaze behaviour and trust ratings in field and simulator. Transp. Res. Part F: Traffic Psychol. Behav. 66, 15–28. http://dx.doi.org/10.1016/j.trf.2019.08.013.

Sun, X., Li, J., Tang, P., Zhou, S., Peng, X., Li, H.N., Wang, Q., 2020. Exploring personalised autonomous vehicles to influence user trust. Cogn. Comput. 12 (6), 1170–1186. http://dx.doi.org/10.1007/s12559-020-09757-x.

Tafidis, P., Farah, H., Brijs, T., Pirdavani, A., 2022. Safety implications of higher levels of automated vehicles: a scoping review. Transp. Rev. 42 (2), 245–267. http://dx.doi.org/10.1080/01441647.2021.1971794.

Tan, X., Zhang, Y., 2025. Driver situation awareness for regaining control from conditionally automated vehicles: a systematic review. Hum. Factors 67 (4), 367–403. http://dx.doi.org/10.1177/00187208241272071.

Taylor, S., Wang, M., Jeon, M., 2023. Reliable and transparent in-vehicle agents lead to higher behavioral trust in conditionally automated driving systems. Front. Psychol. 14, http://dx.doi.org/10.3389/fpsyg.2023.1121622.

Tengilimoglu, O., Carsten, O., Wadud, Z., 2023. Implications of automated vehicles for physical road environment: a comprehensive review. Transp. Res. Part E: Logist. Transp. Rev. 169, 102989. http://dx.doi.org/10.1016/j.tre.2022.102989.

Thomaschke, R., Haering, C., 2014. Predictivity of system delays shortens human response time. Int. J. Hum.–Comput. Stud. 72 (3), 358–365. http://dx.doi.org/10.1016/j.ijhcs.2013.12.004.

Tran, T.T.M., Parker, C., Tomitsch, M., 2021. A review of virtual reality studies on autonomous vehicle–pedestrian interaction. IEEE Trans. Hum.-Mach. Syst. 51 (6), 641–652. http://dx.doi.org/10.1109/THMS.2021.3107517.

Venema, T.A.G., Kroese, F.M., Benjamins, J.S., de Ridder, D.T.D., 2020. When in doubt, follow the crowd? Responsiveness to social proof nudges in the absence of clear preferences. Front. Psychol. 11, 1385. http://dx.doi.org/10.3389/fpsyg.2020.01385.

Walker, F., Forster, Y., Hergeth, S., Kraus, J., Payre, W., Wintersberger, P., Martens, M., 2023. Trust in automated vehicles: constructs, psychological processes, and assessment. Front. Psychol. 14, http://dx.doi.org/10.3389/fpsyg.2023.1279271.

Walker, F., Wang, J., Martens, M.H., Verwey, W.B., 2019. Gaze behaviour and electrodermal activity: objective measures of drivers' trust in automated vehicles. Transp. Res. Part F: Traffic Psychol. Behav. 64, 401–412. http://dx.doi.org/10.1016/j.trf.2019.05.021.

Weinstock, A., Oron-Gilad, T., Parmet, Y., 2012. The effect of system aesthetics on trust, cooperation, satisfaction and annoyance in an imperfect automated system. Work 41 (S1), 258–265. http://dx.doi.org/10.3233/WOR-2012-0166-258.

Wilson, C., Gyi, D., Morris, A., Bateman, H., Tanaka, H., 2022. Non-driving-related tasks and journey types for future autonomous vehicle owners. Transp. Res. Part F: Traffic Psychol. Behav. 85, 150–160. http://dx.doi.org/10.1016/j.trf.2022.01.004.

Yoo, S., Kumagai, J., Managi, S., 2024. Trust more, fear less: the role of social support in fully automated vehicle choice. IEEE Trans. Eng. Manag. 71, 5000–5017. http://dx.doi.org/10.1109/TEM.2022.3208567.

Yoo, J., Kumagai, S., Morita, S., Park, H., Managi, S., 2021. Who to sacrifice? Modeling the driver's dilemma. Transp. Res. Part F: Traffic Psychol. Behav. 78, 308–321.